

WHAT IS CLAIMED IS:

1. A method for determining properties of products from a combinatorial chemical library  $P$  using features of their respective building blocks, the method comprising the steps of:

(1) determining at least one feature for each building block in the combinatorial library  $P$ ,  $\{a_{ijk}, i = 1, 2, \dots, r; j = 1, 2, \dots, r_i; k = 1, 2, \dots, n_i\}$ , wherein  $r$  represents the number of variation sites in the combinatorial library,  $r_i$  represents the number of building blocks at the  $i$ -th variation site, and  $n_i$  represents the number of features used to characterize each building block at the  $i$ -th variation site;

(2) selecting a training subset of products  $\{p_i, i = 1, 2, \dots, m; p_i \in P\}$  from the combinatorial library  $P$ ;

(3) determining  $q$  properties for each compound  $p_i$  in the selected training subset of products, wherein  $y_i = \{y_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, q\}$  represents the determined properties of compound  $p_i$ , and wherein  $q$  is greater or equal to one;

(4) identifying, for each product  $p_i$  of the training subset of products, the corresponding building blocks  $\{t_{ij}, t_{ij} = 1, 2, \dots, r_j, j = 1, 2, \dots, r\}$  and concatenating their features determined in step (1) into a single vector  $\{x_i = a_{1t_{i1}} | a_{2t_{i2}} | \dots | a_{rt_{ir}}\}$ ;

(5) using a supervised machine learning approach to infer a mapping function  $f$  that transforms input values  $x_i$  to output values  $y_i$  from the input/output pairs in the training set  $T = \{(x_i, y_i), i = 1, 2, \dots, m\}$ ;

(6) identifying, after the mapping function  $f$  is determined, for a product  $p_z \in P$ , the corresponding building blocks  $\{t_{zj}, j = 1, 2, \dots, r\}$  and concatenating their features,  $a_{1t_{z1}}, a_{2t_{z2}}, \dots, a_{rt_{zr}}$ , into a single vector  $\{x_z = a_{1t_{z1}} | a_{2t_{z2}} | \dots | a_{rt_{zr}}\}$ , and

(7) mapping  $x_z \rightarrow y_z$ , using the mapping function  $f$  determined in step (5), wherein  $y_z$  represents the properties of product  $p_z$ .

2. The method of claim 1, wherein step (1) comprises the step of: using a measured value as a feature for each building block.
3. The method of claim 1, wherein step (1) comprises the step of: using a computed value as a feature for each building block.
4. The method of claim 1, wherein step (3) comprises the step of: using a measured value as a property for each product of the training subset.
5. The method of claim 1, wherein step (3) comprises the step of: using a computed value as a property for each product of the training subset.
6. The method of claim 1, wherein step (5) comprises the step of: training a multilayer perceptron.
7. The method of claim 1, wherein at least one of the features determined in step (1) is the same as at least one of the properties determined in step (3).
8. The method of claim 1, wherein the building blocks comprise a plurality of reagents used to construct the combinatorial library  $P$ .
9. The method of claim 1, wherein the building blocks comprise a plurality of fragments of a plurality of reagents used to construct the combinatorial library  $P$ .
10. The method of claim 1, wherein the building blocks comprise a plurality of modified fragments of a plurality of reagents used to construct the combinatorial library  $P$ .

11. The method of claim 1, wherein step (2) comprises the step of: selecting a training subset of products at random.
12. The method of claim 1, wherein step (2) comprises the step of: selecting a training subset of products using a combinatorial design method to cover all pairwise combinations of building blocks.
13. The method of claim 1, wherein step (2) comprises the step of: selecting a training subset of products using a diversity metric to select a diverse subset of products.
14. A method for determining properties of combinatorial library products from features of library building blocks, the method comprising the steps of:
  - (1) determining at least one feature for each building block of a combinatorial library having a plurality of products;
  - (2) selecting a training subset of products from the plurality of products of the combinatorial library;
  - (3) determining at least one property for each product of the training subset of products;
  - (4) identifying a building block set for each product of the training subset of products;
  - (5) forming an input features vector for each product of the training subset of products from the building block set for each product of the training subset of products;
  - (6) using a supervised machine learning approach to infer a mapping function  $f$  that transforms the input features vector for each product of the training subset of products to the corresponding at least one property for each product of the training subset of products;
  - (7) identifying building block sets for a plurality of additional products of the combinatorial library;

(8) forming input features vectors for the plurality of additional products from the building block sets for the plurality of additional products; and

(9) transforming the input features vectors for the plurality of additional products using the mapping function  $f$  to obtain at least one estimate property for each of the plurality of additional products.

15. The method of claim 14, wherein step (1) comprises the step of: using a measured value as a feature for each building block of the combinatorial library.
16. The method of claim 14, wherein step (1) comprises the step of: using a computed value as a feature for each building block of the combinatorial library.
17. The method of claim 14, wherein step (3) comprises the step of: using a measured value as a property for each product of the training subset of products.
18. The method of claim 14, wherein step (3) comprises the step of: using a computed value as a property for each product of the training subset of products.
19. The method of claim 14, wherein step (6) comprises the step of: training a multilayer perceptron using the input features vector and the corresponding at least one property for each product of the training subset of products.
20. The method of claim 14, wherein at least one of the features determined in step (1) is the same as at least one of the properties determined in step (3).

21. The method of claim 14, wherein  
the building blocks of the combinatorial library comprise a plurality of reagents used to construct the combinatorial library.
  
22. The method of claim 14, wherein  
the building blocks of the combinatorial library comprise a plurality of fragments of a plurality of reagents used to construct the combinatorial library.
  
23. The method of claim 14, wherein  
the building blocks of the combinatorial library comprise a plurality of modified fragments of a plurality of reagents used to construct the combinatorial library.
  
24. The method of claim 14, wherein step (2) comprises the step of:  
selecting a training subset of products at random.
  
25. The method of claim 14, wherein step (2) comprises the step of:  
selecting a training subset of products using a combinatorial design method to cover all pairwise combinations of building blocks.
  
26. The method of claim 14, wherein step (2) comprises the step of:  
selecting a training subset of products using a diversity metric to select a diverse subset of products.
  
27. A system for determining properties of combinatorial library products from features of library building blocks, comprising:  
a module for determining at least one feature for each building block of a combinatorial library having a plurality of products;  
a module for selecting a training subset of products from the plurality of products of the combinatorial library;  
a module for determining at least one property for each product of the training subset of products;

a module for identifying a building block set for each product of the training subset of products;

a module for forming an input features vector for each product of the training subset of products from the building block set for each product of the training subset of products;

a module for using a supervised machine learning approach to infer a mapping function  $f$  that transforms the input features vector for each product of the training subset of products to the corresponding at least one property for each product of the training subset of products;

a module for identifying building block sets for a plurality of additional products of the combinatorial library;

a module for forming input features vectors for the plurality of additional products from the building block sets for the plurality of additional products; and

a module for transforming the input features vectors for the plurality of additional products using the mapping function  $f$  to obtain at least one estimate property for each of the plurality of additional products.

28. A system for determining properties of combinatorial library products from features of library building blocks, comprising:

means for determining at least one feature for each building block of a combinatorial library having a plurality of products;

means for selecting a training subset of products from the plurality of products of the combinatorial library;

means for determining at least one property for each product of the training subset of products;

means for identifying a building block set for each product of the training subset of products;

means for forming an input features vector for each product of the training subset of products from the building block set for each product of the training subset of products;

means for using a supervised machine learning approach to infer a mapping function  $f$  that transforms the input features vector for each

product of the training subset of products to the corresponding at least one property for each product of the training subset of products;

means for identifying building block sets for a plurality of additional products of the combinatorial library;

means for forming input features vectors for the plurality of additional products from the building block sets for the plurality of additional products; and

means for transforming the input features vectors for the plurality of additional products using the mapping function  $f$  to obtain at least one estimate property for each of the plurality of additional products.

29. A computer program product for determining properties of combinatorial library products from features of library building blocks, said computer program product comprising a computer useable medium having computer program logic recorded thereon for controlling a processor, said computer program logic comprising:

a procedure that enables said processor to determine at least one feature for each building block of a combinatorial library having a plurality of products;

a procedure that enables said processor to select a training subset of products from the plurality of products of the combinatorial library;

a procedure that enables said processor to determine at least one property for each product of the training subset of products;

a procedure that enables said processor to identify a building block set for each product of the training subset of products;

a procedure that enables said processor to form an input features vector for each product of the training subset of products from the building block set for each product of the training subset of products;

a procedure that enables said processor to use a supervised machine learning approach to infer a mapping function  $f$  that transforms the input features vector for each product of the training subset of products to the corresponding at least one property for each product of the training subset of products;

a procedure that enables said processor to identify building block sets for a plurality of additional products of the combinatorial library;

a procedure that enables said processor to form input features vectors for the plurality of additional products from the building block sets for the plurality of additional products; and

a procedure that enables said processor to transform the input features vectors for the plurality of additional products using the mapping function  $f$  to obtain at least one estimate property for each of the plurality of additional products.

30. The computer program product of claim 29, further comprising:

a procedure that enables said processor to train a multilayer perceptron using the input features vector and the corresponding at least one property for each product of the training subset of products.

31. The computer program product of claim 29, further comprising:

a procedure that enables said processor to use a measured value as a property for each product of the training subset of products.

32. The computer program product of claim 29, further comprising:

a procedure that enables said processor to use a computed value as a property for each product of the training subset of products.

33. The computer program product of claim 29, further comprising:

a procedure that enables said processor to use a measured value as a feature for each building block of the combinatorial library.

34. The computer program product of claim 29, further comprising:

a procedure that enables said processor to use a computed value as a feature for each building block of the combinatorial library.

35. The computer program product of claim 29, wherein the building blocks of the combinatorial library comprise a plurality of reagents used to construct the combinatorial library.
36. The computer program product of claim 29, wherein the building blocks of the combinatorial library comprise a plurality of fragments of a plurality of reagents used to construct the combinatorial library.
37. The computer program product of claim 29, wherein the building blocks of the combinatorial library comprise a plurality of modified fragments of a plurality of reagents used to construct the combinatorial library.
38. The computer program product of claim 29, further comprising: a procedure that enables said processor to select the training subset of products at random.
39. The computer program product of claim 29, further comprising: a procedure that enables said processor to select the training subset of products using a combinatorial design method to cover all pairwise combinations of building blocks.
40. The computer program product of claim 29, further comprising: a procedure that enables said processor to select the training subset of products using a diversity metric to select a diverse subset of products.